Earth Science Markup Language (ESML)

Ken Keiser, Helen Conover, Rahul Ramachandran, Sara Graves

Information Technology and Systems Center

University of Alabama in Huntsville
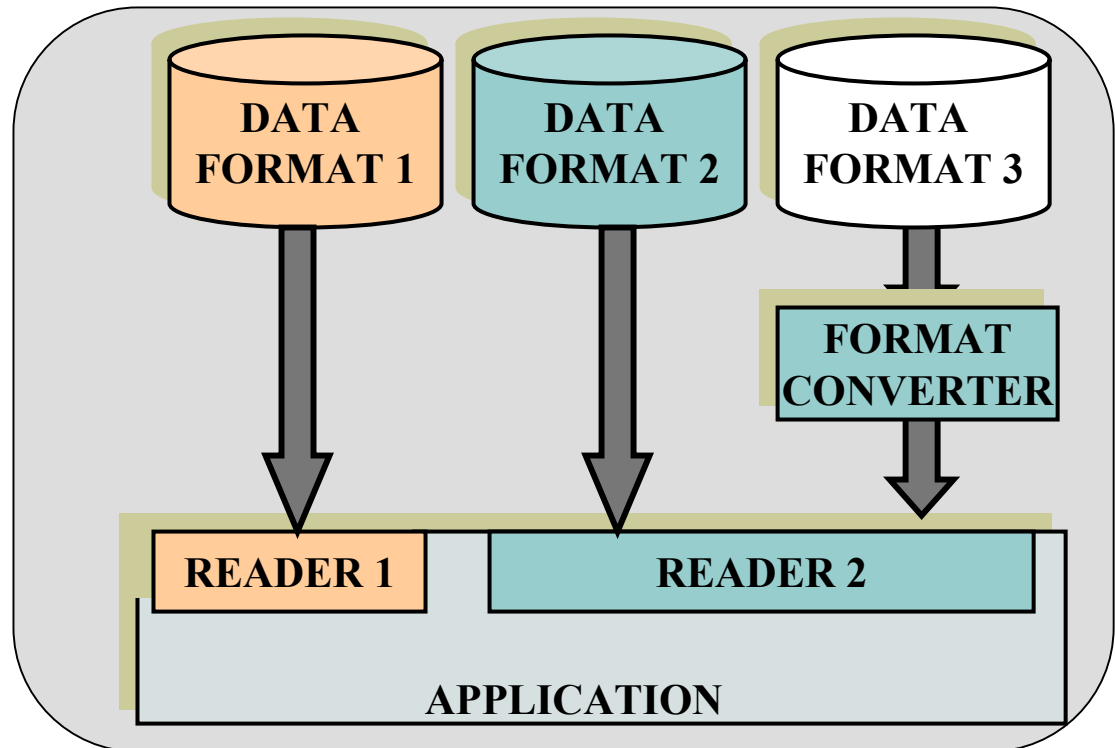
esml

RFC Content

- Standard or Technical Note?
  - Metadata
  - Interface specification
- Motivation for RFC?
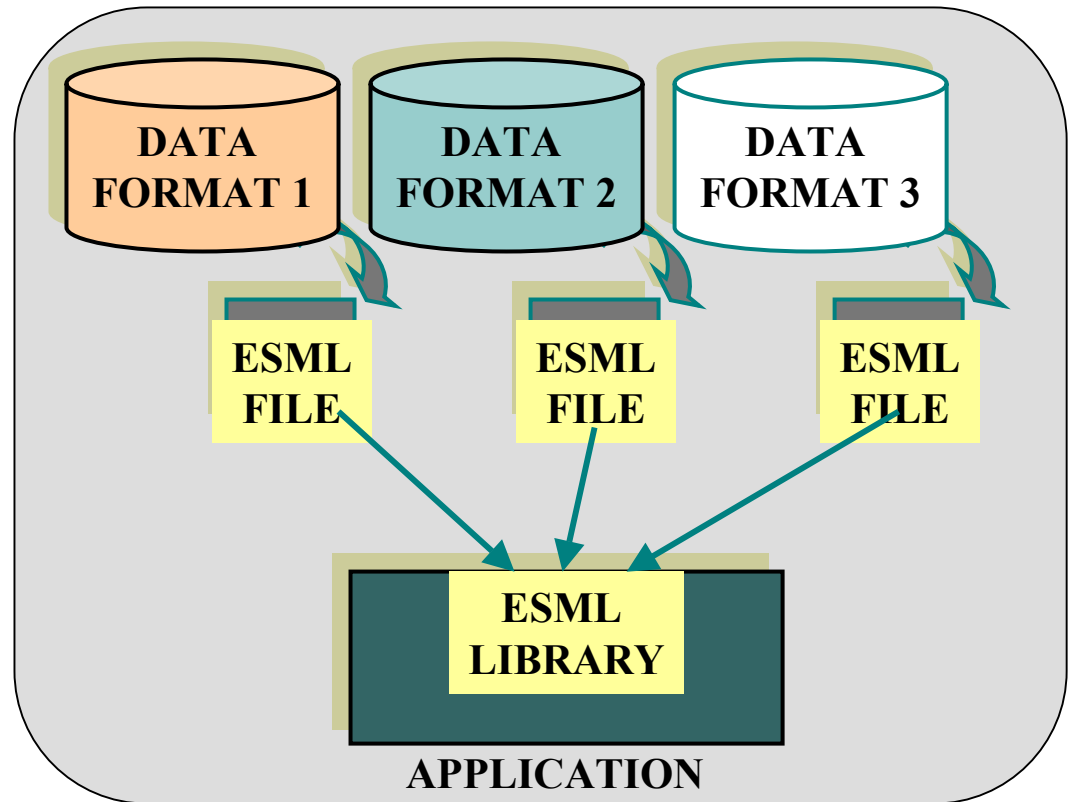- Overview of Technology

esml

Data Usability Problem

- Specialized code is required for every format
  - Difficult to assimilate new data types
  - Makes applications tightly coupled to data
- Standardized formats not widely accepted and requires conversion of legacy data
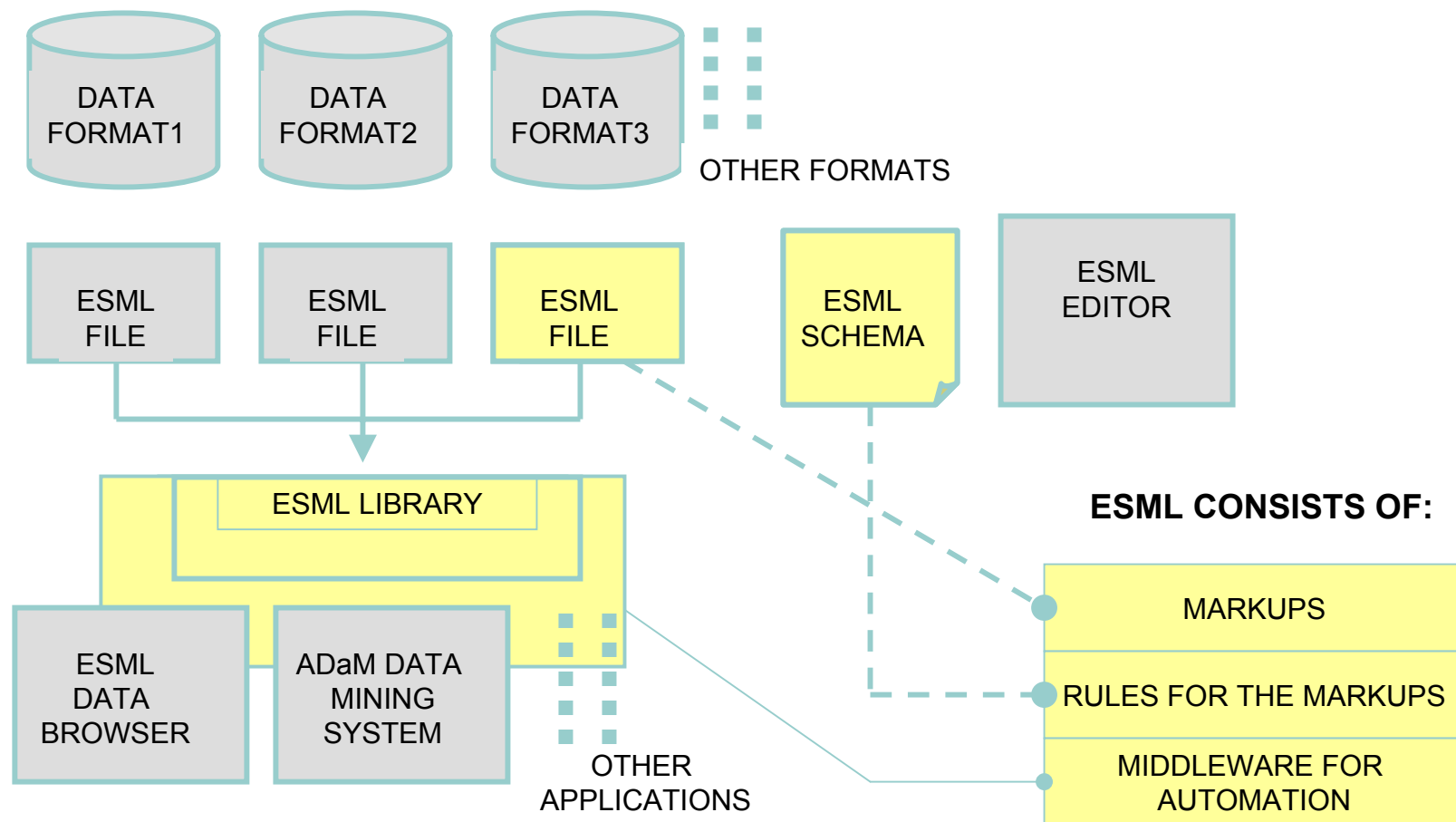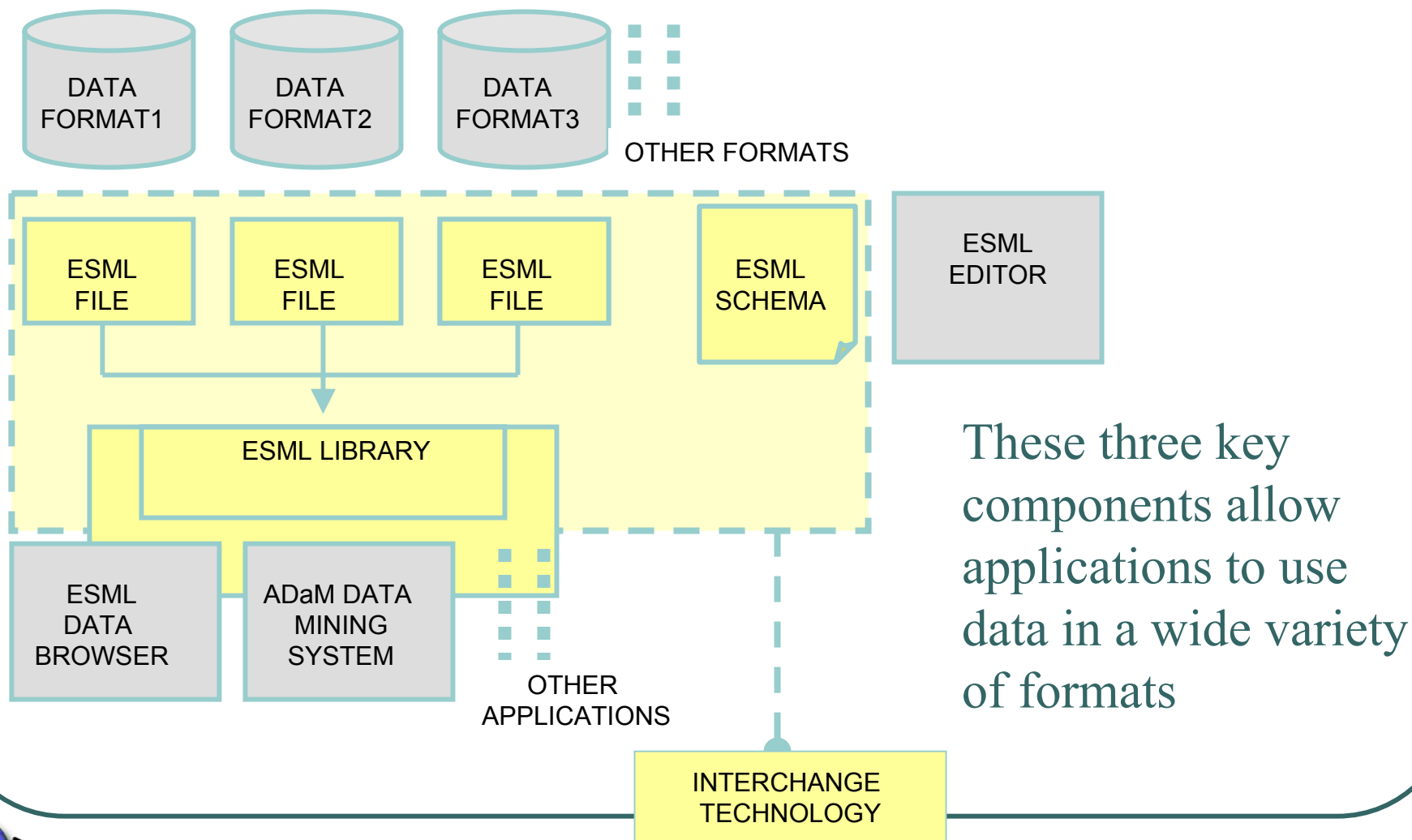
ESML Solution

- ESML (external metadata) files containing the structural description of the data format

- Applications utilize these descriptions to figure out how to read the data files resulting in data interoperability for applications

| DATA FORMAT 1 | DATA FORMAT 2 | DATA FORMAT 3 |

ESML FILE   ESML FILE   ESML FILE

ESML LIBRARY

APPLICATION

esml

Components of the ESML Interchange Technology

DATA FORMAT1

DATA FORMAT2

DATA FORMAT3

OTHER FORMATS

ESML FILE

ESML FILE

ESML FILE

ESML SCHEMA

ESML EDITOR

ESML LIBRARY

ESML CONSISTS OF:

ESML DATA BROWSER

ADaM DATA MINING SYSTEM

OTHER APPLICATIONS

MARKUPS

RULES FOR THE MARKUPS

MIDDLEWARE FOR AUTOMATION

esml

Components of the ESML Interchange Technology

DATA FORMAT1

DATA FORMAT2

DATA FORMAT3

OTHER FORMATS

ESML FILE

ESML FILE

ESML FILE

ESML SCHEMA

ESML EDITOR

ESML LIBRARY

ESML DATA BROWSER

ADaM DATA MINING SYSTEM

OTHER APPLICATIONS

INTERCHANGE TECHNOLOGY

These three key components allow applications to use data in a wide variety of formats

esml

Interchange Technology for Data Users and Application Developers

DATA FORMAT1

DATA FORMAT2

DATA FORMAT3

OTHER FORMATS

DATA PRODUCERS OR CONSUMERS

ESML FILE

ESML FILE

ESML FILE

ESML SCHEMA

ESML EDITOR

ESML LIBRARY

ESML DATA BROWSER

ADaM DATA MINING SYSTEM

OTHER APPLICATIONS

ESML can be used by both scientists and application developers

APPLICATION DEVELOPERS

INTERCHANGE TECHNOLOGY
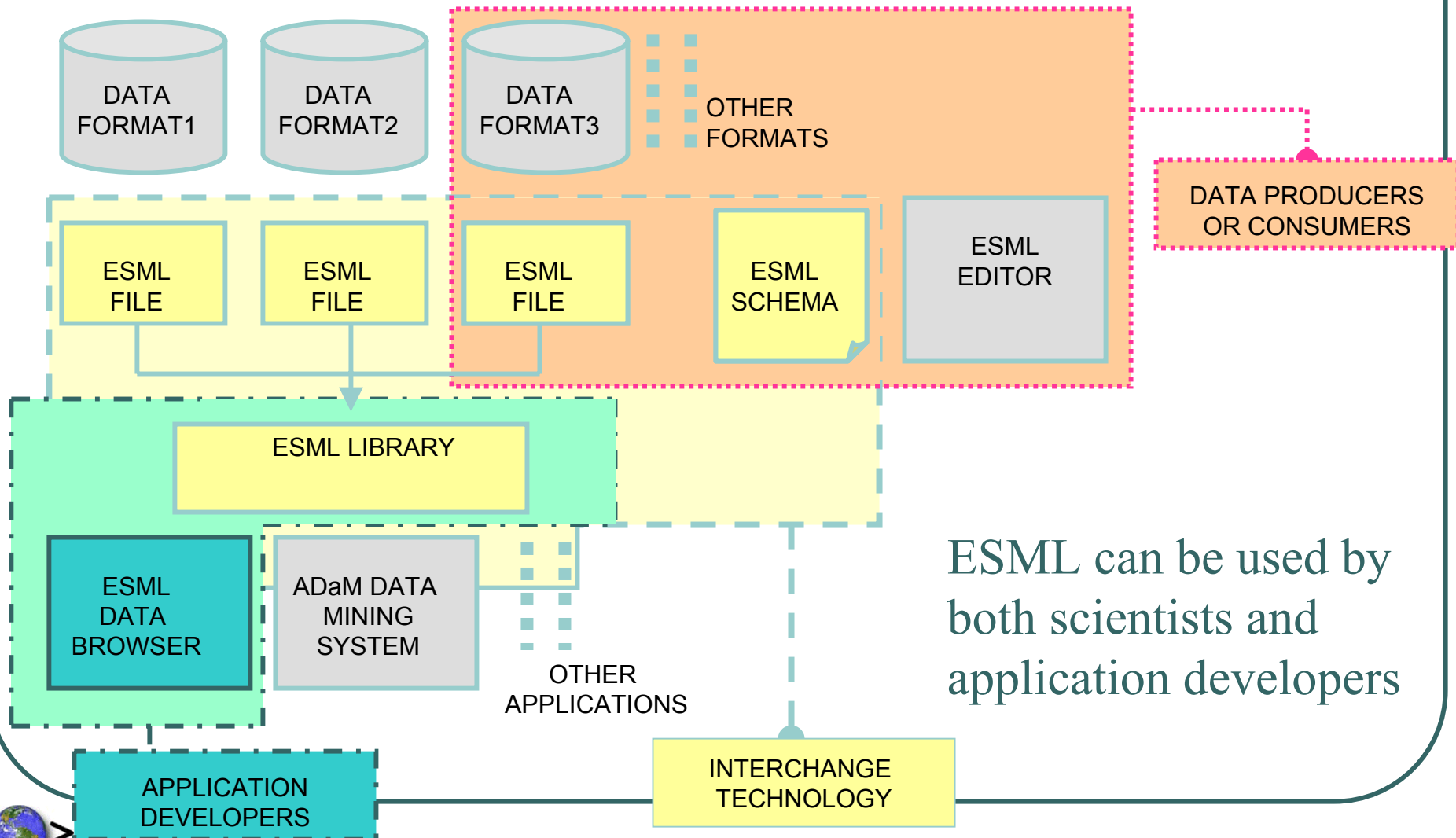
esml

ESML Components
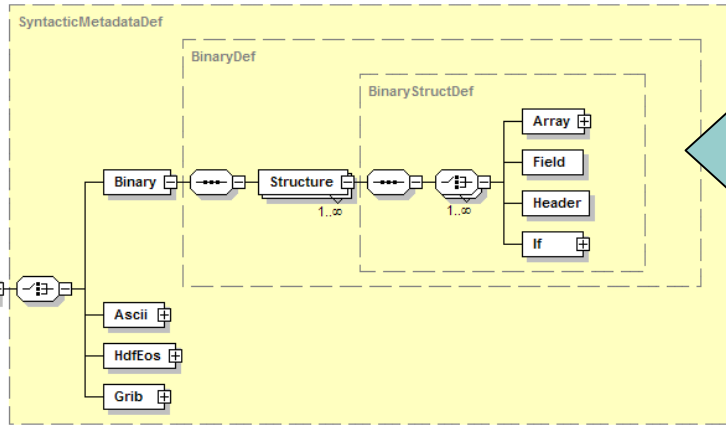
ESML Schema defines syntactic metadata that describe the structure of the file in machine-readable and interpretable terms and are the grammar that are used to generate the ESML description file

ESML Description Files specify the structure of the data file format in terms of bits and bytes

ESML Library is the middleware that applications use to parse an ESML Description File and retrieve data
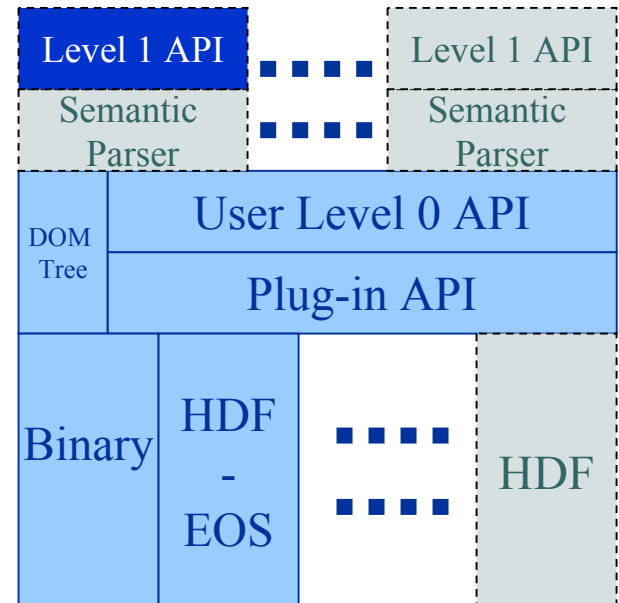
ESML MARKUP FOR THE DATA FILE

```
<a:ESML >
 <SyntacticMetaData>
  <Ascii>
   <Structure instances="1">
    <Header name="SizeX" format="%d" symbol="true" />
    <Header name="SizeY" format="%d" symbol="true" />
    <Array occurs="$SizeX">
     <Array occurs="$SizeY">
      <Field name="BrightnessTemp" format="%d"/>
     </Array>
    </Array>
   </Structure>
  </Ascii>
 </SyntacticMetaData>
</a:ESML>
```

SIMPLE ASCII DATA FILE

```
4
5
 1  2  3  4  5
 6  7  8  9 10
11 12 13 14 15
16 17 18 19 20
```

ESML v3.0 Library Layered Design

Level 1 API	· · · · ·	Level 1 API	
Semantic Parser	· · · · ·	Semantic Parser	

DOM Tree	User Level 0 API	
	Plug-in API	

Binary	HDF - EOS	· · · · ·	HDF
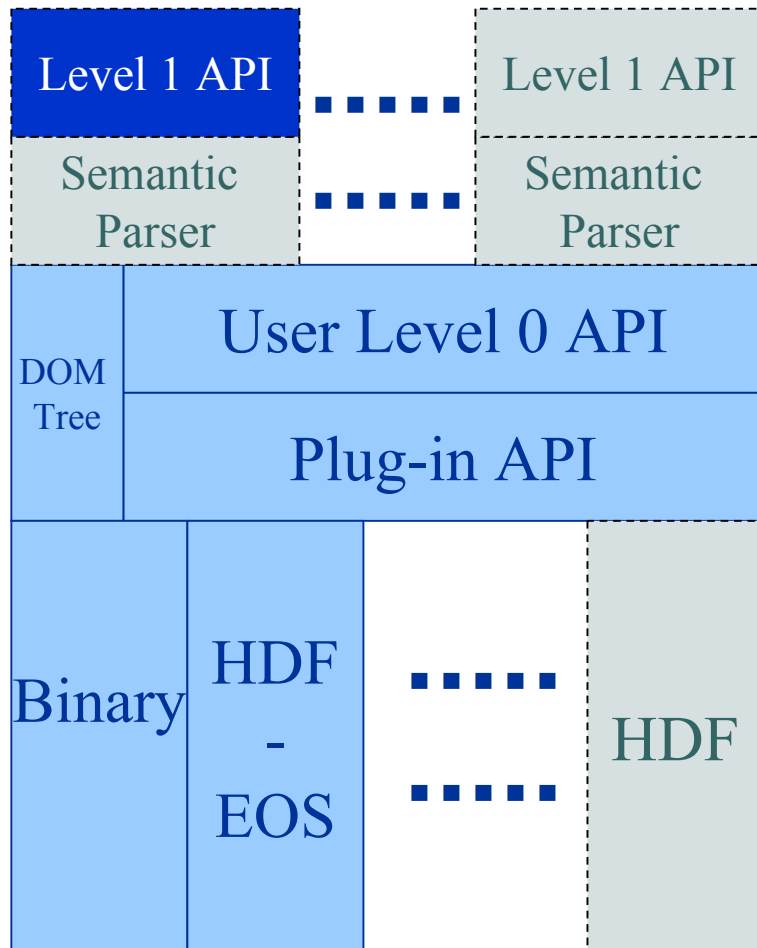
The **core ESML library** provides the basic functionality of reading structural metadata from the ESML file and returning data to the user

- Intuitive **user API** based on the analogy of file access in a directory structure

- **Plug-in modules** for each individual format allow flexible packaging of libraries

- Simple **Plug-in API** for easy addition of new formats

- **Additional software** can be easily added to provide other functions such as using semantics from an ontology to "use" the data intelligently

esml

```
PythonWin                                                          _ □ ×
File  Edit  View  Tools  Window  Help

[toolbar icons]

Interactive Window                                                 _ □ ×
>>> import pyESML
>>> dir(pyESML)
['__doc__', '__file__', '__name__', 'getRank', 'isField', 'isStructure', 'list', 'listF
>>> obj = pyESML.open('SimpleBinary.xml','TestBin.dat')
>>> print pyESML.listStructures(obj)
['Bin']
>>> print pyESML.listFields(obj)
['/Bin/BrightnessTemp']
>>> data = pyESML.readField(obj,'/Bin/BrightnessTemp')
>>> for val in data:
...     print val
...
0.0
1.0
2.0
3.0
4.0
5.0
6.0
7.0
Ready                                                    00042  005
```

List all the
API functions

Open a data
file using ESML

List all the
structures

List all the
fields

Get the data
for a field

Print the data

esml

What is ESML?

- It is a **specialized markup language** for Earth Science metadata based on XML - NOT another data format.

- It is a machine-readable and -interpretable representation of the **structure, semantics and content** of any data file, regardless of data format

- ESML description files contain **external metadata** that can be generated by either data producer or data consumer (at collection, data set, and/or granule level)

- ESML provides the **benefits of a standard, self-describing data format** (like HDF, HDF-EOS, netCDF, geoTIFF, …) without the cost of data conversion

- ESML **complements and extends data catalogs** such as FGDC and GCMD by providing the use/access information those directories lack.
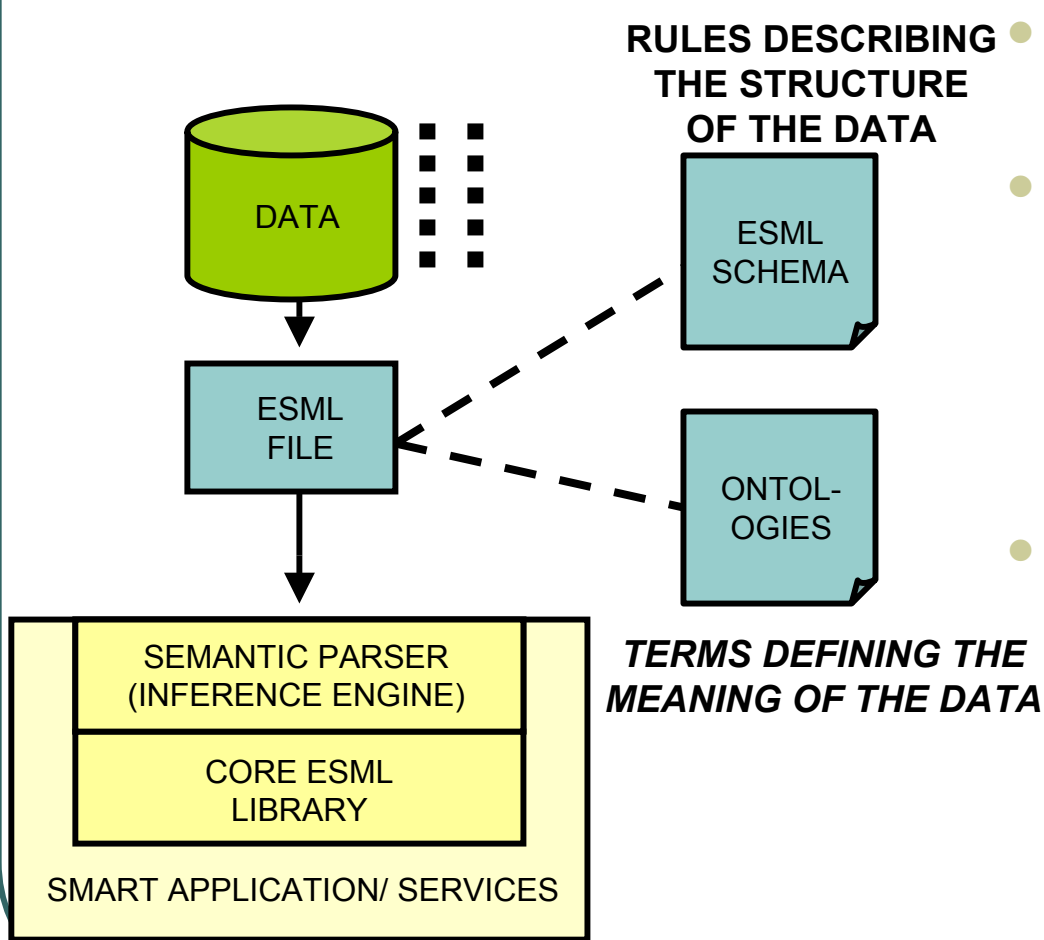
Potential Benefits of ESML

- Software access to heterogeneous data formats
- Software insulation to format changes (change the description not the code)
- Single description for file groups that are syntactically and semantically similar
- Provides access solution for legacy data sets
- Data provider or user can generate/provide descriptions for requested data sets
- Can be used to provide multiple externalized views of same data

How might we add semantics?
Example: Extending ESML with Ontologies

RULES DESCRIBING THE STRUCTURE OF THE DATA

DATA

ESML FILE

ESML SCHEMA

ONTOL-OGIES

SEMANTIC PARSER (INFERENCE ENGINE)

CORE ESML LIBRARY

SMART APPLICATION/ SERVICES

TERMS DEFINING THE MEANING OF THE DATA

- ESML Schema provides structural metadata

- Extend ESML schema by embedding semantic terms in the ESML Description File to provide a complete description of the data

- Allow various science communities to create their own ontologies (for example, SWEET) and use them with ESML Description Files for their data

User Communities

- Science researchers
- Application developers
- Data providers (ESML descriptions)
  - Data producers
  - Data archives
- 54 Registered Users (US edu, mil, gov, com: international edu, gov)

esml

Current Status

- Open Source (version 3.0.2)
  - Avg 38 downloads/month since Aug 2003 (first version 3 release)
- ESML data formats
  - Currently supported: ASCII, Binary, HDF-EOS, HDF-5, NetCDF, Grib, NEXRAD Level II
  - In work: HDF-4
- ESML Library
  - C++ for Windows and Linux
  - Python API supported, FORTRAN API in work
  - OPeNDAP / ESML server
  - IDL Plug-in

http://esml.itsc.uah.edu

Recent Inquiries & Interactions

- Known NASA Project Involvement
  - GPM (FORTRAN wrappers, analysis tool plug-ins)
    - Erich Stocker (NASA/GSFC)
  - Terrestrial Observation and Prediction System (TOPS)
    - Petr Votava (NASA/Ames)
  - Langley DAAC
    - Inquiries from Linda Hunt
- Other Interactions
  - TVA Environmental Technologies
    - Cary West: communications about "the use of ESML tools in the analysis of our MM5/CMAQ model output"
  - ESA Earth Observation Programme
    - Silvia de Castro: "applying it to binary data files, products of Earth Observation satellites"
  - Global Grid Forum
    - ESML designers contributing to similar Grid-oriented Data Format Definition Language (DFDL)

esml